



**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**

BRNO UNIVERSITY OF TECHNOLOGY

**FAKULTA INFORMAČNÍCH TECHNOLOGIÍ**

FACULTY OF INFORMATION TECHNOLOGY

**ÚSTAV POČÍTAČOVÉ GRAFIKY A MULTIMÉDIÍ**

DEPARTMENT OF COMPUTER GRAPHICS AND MULTIMEDIA

**ROZPOZNÁVÁNÍ HISTORICKÝCH TEXTŮ POMOCÍ HLU-  
BOKÝCH NEURONOVÝCH SÍTÍ**

CONVOLUTIONAL NETWORKS FOR HISTORIC TEXT RECOGNITION

**SEMESTRÁLNÍ PROJEKT**

TERM PROJECT

**AUTOR PRÁCE**

AUTHOR

**Bc. MARTIN KIŠŠ**

**VEDOUCÍ PRÁCE**

SUPERVISOR

**Ing. MICHAL HRADIŠ, Ph.D.**

**BRNO 2018**

## Abstrakt

Do tohoto odstavce bude zapsán výtah (abstrakt) práce v českém (slovenském) jazyce.

## Abstract

Do tohoto odstavce bude zapsán výtah (abstrakt) práce v anglickém jazyce.

## Klíčová slova

Sem budou zapsána jednotlivá klíčová slova v českém (slovenském) jazyce, oddělená čárkami.

## Keywords

Sem budou zapsána jednotlivá klíčová slova v anglickém jazyce, oddělená čárkami.

## Citace

KIŠŠ, Martin. *Rozpoznávání historických textů pomocí hlubokých neuronových sítí*. Brno, 2018. Semestrální projekt. Vysoké učení technické v Brně, Fakulta informačních technologií. Vedoucí práce Ing. Michal Hradiš, Ph.D.

# Rozpoznávání historických textů pomocí hlubokých neuronových sítí

## Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně pod vedením pana X... Další informace mi poskytli... Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

.....

Martin Kišš  
20. prosince 2017

## Poděkování

V této sekci je možno uvést poděkování vedoucímu práce a těm, kteří poskytli odbornou pomoc (externí zadavatel, konzultant, apod.).

# Obsah

<b>1</b>	<b>Úvod</b>	<b>2</b>
<b>2</b>	<b>Neuronové sítě</b>	<b>3</b>
2.1	Konvoluční neuronové sítě . . . . .	3
2.2	Rekurentní neuronové sítě . . . . .	3
2.3	Spatial Transformer Network . . . . .	3
<b>3</b>	<b>Rozpoznávání textu</b>	<b>4</b>
3.1	Současné metody založené na neuronových sítích . . . . .	4
3.2	Historická písma . . . . .	4
<b>4</b>	<b>Cíl práce</b>	<b>5</b>
<b>5</b>	<b>Datové sady?</b>	<b>6</b>
<b>6</b>	<b>Stav práce</b>	<b>7</b>
6.1	Generátor textů . . . . .	7
6.1.1	Efekty . . . . .	8
6.2	Klasifikace umělých textů . . . . .	8
6.2.1	Datová sada . . . . .	8
6.2.2	Výsledky experimentu . . . . .	8
<b>7</b>	<b>Závěr</b>	<b>9</b>
	<b>Literatura</b>	<b>10</b>

# Kapitola 1

## Úvod

Rozpoznávání znaků (anglicky Optical Character Recognition, zkráceně OCR), respektive celých textů, je proces, ve kterém se převádí text z obrazové formy do sekvence znaků, které mohou být dále zpracovány. Vzhledem ke stále větší integraci počítačových systémů do běžného života, může tato úloha nacházet široké uplatnění. Například se může jednat o součást systému řízení autonomního vozidla, které díky rozpoznávání textů je schopné číst dopravní značení. Dále může být součástí aplikací zabývajících se překladem z různých jazyků, kdy se překládaný text zadává pomocí kamery (například v chytrém mobilním telefonu). Využití je také možné při vyhledávání obrázků, kdy se díky rozpoznání textů může získat z obrázku více informací, a podobně.

V rámci počítačového vidění se jedná o problém, který je spojen s různými úskalími, která mohou převod z obrazové formy ztížit. Jedná se především o data (např. fotografie), která jsou zašuměná, v různých částech obrazu obsahují různou intenzitu světla, nebo například text může být natočen, zkreslen perspektivou a tak dále.

Cílem této práce je vytvořit nástroj, který bude schopen rozpoznávat text z dokumentů, které vznikly v dřívějších dobách. Jedná se především o text z období středověku, kdy se objevují texty psané ručně a také texty vytvořené pomocí prvních metod knihtisku.

V následující kapitole budou popsány konvoluční neuronové sítě, porovnání s rekurentními neuronovými sítě a jejich specifika. V kapitole 3 budou popsány techniky, které se v současnosti používají k rozpoznávání textu. Dalším obsahem této kapitoly také bude stručný popis vývoje historických písem. Kapitola 4 bude obsahovat zvolenou metodu, která bude v rámci práce implementována a následně vyhodnocena. Další kapitola se bude věnovat popisu datových sad, které jsou pro tuto práci vhodné na otestování. V předposlední kapitole bude popsán současný stav práce a poslední kapitola bude obsahovat závěrečné shrnutí.

## Kapitola 2

# Neuronové sítě

### 2.1 Konvoluční neuronové sítě

Základny CNN.

### 2.2 Rekurentní neuronové sítě

Rozdíl mezi CNN a RNN, použití CRNN při zpracování sekvencí, LSTM, GRU.

### 2.3 Spatial Transformer Network

Popis STN a použití. [\[1\]](#)

## Kapitola 3

# Rozpoznávání textu

### 3.1 Současné metody založené na neuronových sítích

Informace čerpané ze zdrojů ohledně OCR (techniky, postupy, ...) [5, 2, 4, 3] Tesseract?

### 3.2 Historická písma

Stručný přehled vývoje historického písma.

## Kapitola 4

# Cíl práce

CRNN, STN.



## Kapitola 5

# Datové sady?

Informace ohledně existujících datových sad, jejich stručný popis, atd.

## Kapitola 6

# Stav práce

V této kapitole bude popsán stav dosavadní práce. Nejprve zde bude popsán implementovaný generátor historických textů. Motivací k jeho vytvoření bylo, že se nepodařilo najít vhodnou datovou sadu, která by obsahovala naskenované historické texty a jejich přepis (tzv. ground truth). Výhodou takto implementovaného generátoru je také množství dat, která mohou být vygenerována, a spolu s tím také možnost přesné lokalizace všech písmen. Ve druhé části této kapitoly bude popsána vygenerovaná datová sada a neuronová síť, která byla na této sadě natrénována. Budou zde také zhodnoceny výsledky, které síť dosáhla.

### 6.1 Generátor textů

Součástí této práce je generátor umělých textů. Tento generátor vytváří syntetické obrázky takové, aby co nejvíce vypadaly jako originální texty naskenované z historických dokumentů. Ke generování obrázků, které vypadají co nejvěrohodněji, je zapotřebí několika zdrojů. Prvním zdrojem jsou bezpochyby vhodné fonty, které v maximální možné míře odpovídají dobovým písmům. Druhým zdrojem jsou obrázky (textury), které mají vzhled starého papíru a na něž budou výsledné vygenerované texty naneseny.

Aby vygenerované obrázky co nejvíce odpovídaly reálným textům, je na původní vysázený text aplikováno několik takzvaných efektů. Tyto efekty mají za úkol upravovat originální vysázený text tak, aby se co nejvíce přiblížil vzhledu historických textů. Tyto efekty budou popsány v následující části **6.1.1**.

K tomu, aby efekty neupravovaly vstupní obrázek pokaždé stejným způsobem, je v generátoru několikrát použito generování náhodných čísel. Aby se daly jednoduše ovládat minimální a maximální hodnoty při tomto generování náhodných čísel, je jediným parametrem při spuštění konfigurační soubor, který obsahuje mimo jiné právě tyto minimální a maximální hodnoty pro jednotlivé efekty. Dalšími parametry, které jsou uloženy v rámci konfiguračního souboru, jsou například rozměry výsledného obrázku, adresáře s fonty a podobně. Ukázkou celého konfiguračního souboru je možné nalézt v příloze.

K vysázení daného textu do obrázku byla použita knihovna freetype ve verzi pro python <sup>1</sup>. Tato knihovna umožňuje vysázení vlastního textu na poměrně nízké úrovni a je tedy možné během sázení uchovávat užitečné informace, jako například již zmíněné pozice jednotlivých znaků.

---

<sup>1</sup><https://pypi.python.org/pypi/freetype-py/>

### **6.1.1 Efekty**

## **6.2 Klasifikace umělých textů**

### **6.2.1 Datová sada**

Popis vygenerované datové sady.

### **6.2.2 Výsledky experimentu**

Popis experimentu ohledně klasifikace textů vytvořených generátorem Zhodnocení na reálné sadě?

## Kapitola 7

# Závěr

Shrnutí a nastínění další práce.

# Literatura

- [1] Jaderberg, M.; Simonyan, K.; Zisserman, A.; aj.: *Spatial Transformer Network*. Únor 2016, [Online; navštíveno 19.12.2017].  
URL <https://arxiv.org/abs/1506.02025>
- [2] Shi, B.; Bai, X.; Yao, C.: *An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition*. Prosinec 2016, [Online; navštíveno 19.12.2017].  
URL <http://ieeexplore.ieee.org/abstract/document/7801919/>
- [3] Shi, B.; Wang, X.; Lyu, P.; aj.: *Robust Scene Text Recognition with Automatic Rectification*. 2016, [Online; navštíveno 19.12.2017].  
URL [https://www.cv-foundation.org/openaccess/content\\_cvpr\\_2016/html/Shi\\_Robust\\_Scene\\_Text\\_CVPR\\_2016\\_paper.html](https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Shi_Robust_Scene_Text_CVPR_2016_paper.html)
- [4] Springmann, U.; Lüdeling, A.: *OCR of historical printings with an application to building diachronic corpora: A case study using the RIDGES herbal corpus*. Únor 2017, [Online; navštíveno 19.12.2017].  
URL <https://arxiv.org/abs/1608.02153>
- [5] Su, B.; Lu, S.: *Accurate recognition of words in scenes without character segmentation using recurrent neural network*. Říjen 2016, [Online; navštíveno 19.12.2017].  
URL <http://www.sciencedirect.com/science/article/pii/S0031320316303314>